

Laravel AI SDK

Introduction

Installation

Configuration

Custom Base URLs

Provider Support

Agents

Prompting

Conversation Context

Structured Output

Attachments

Streaming

Broadcasting

Queueing

Tools

Provider Tools

Middleware

Anonymous Agents

Agent Configuration

Images

Audio (TTS)

Transcription (STT)

Embeddings

Querying Embeddings

Caching Embeddings

Reranking

Files

Vector Stores

Adding Files to Stores

Failover

Testing

Agents

Images

Audio

Transcriptions

Embeddings

Reranking

Files

Vector Stores

Events

Introduction

The [Laravel AI SDK](#) provides a unified, expressive API for interacting with AI providers such as OpenAI, Anthropic, Gemini, and more. With the AI SDK, you can build intelligent agents with tools and structured output, generate images, synthesize and transcribe audio, create vector embeddings, and much more — all using a consistent, Laravel-friendly interface.

Installation

You can install the Laravel AI SDK via Composer:

```
1 composer require laravel/ai
```

Next, you should publish the AI SDK configuration and migration files using the `vendor:publish` Artisan command:

```
1 php artisan vendor:publish --provider="Laravel\Ai\AiServiceProvider"
```

Finally, you should run your application's database migrations. This will create a `agent_conversations` and `agent_conversation_messages` table that the AI SDK uses to power its conversation storage:

```
1 php artisan migrate
```

Configuration

You may define your AI provider credentials in your application's `config/ai.php` configuration file or as environment variables in your application's `.env` file:

```
1 ANTHROPIC_API_KEY=
2 COHERE_API_KEY=
3 ELEVENLABS_API_KEY=
4 GEMINI_API_KEY=
5 MISTRAL_API_KEY=
6 OLLAMA_API_KEY=
7 OPENAI_API_KEY=
8 JINA_API_KEY=
9 VOYAGEAI_API_KEY=
10 XAI_API_KEY=
```

The default models used for text, images, audio, transcription, and embeddings may also be configured in your application's `config/ai.php` configuration file.

Custom Base URLs

By default, the Laravel AI SDK connects directly to each provider's public API endpoint. However, you may need to route requests through a different endpoint - for example, when using a proxy service to centralize API key management, implement rate limiting, or route traffic through a corporate gateway.

You may configure custom base URLs by adding a `url` parameter to your provider configuration:

```
1 'providers' => [
2     'openai' => [
3         'driver' => 'openai',
4         'key' => env('OPENAI_API_KEY'),
5         'url' => env('OPENAI_BASE_URL'),
6     ],
7     'anthropic' => [
8         'driver' => 'anthropic',
9         'key' => env('ANTHROPIC_API_KEY'),
10        'url' => env('ANTHROPIC_BASE_URL'),
11    ],
12 ],
13 ]
```

This is useful when routing requests through a proxy service (such as LiteLLM or Azure OpenAI Gateway) or using alternative endpoints.

Custom base URLs are supported for the following providers: OpenAI, Anthropic, Gemini, Groq, Cohere, DeepSeek, xAI, and OpenRouter.

Provider Support

The AI SDK supports a variety of providers across its features. The following table summarizes which providers are available for each feature:

Feature	Providers
Text	OpenAI, Anthropic, Gemini, Azure, Groq, xAI, DeepSeek, Mistral, Ollama
Images	OpenAI, Gemini, xAI
TTS	OpenAI, ElevenLabs
STT	OpenAI, ElevenLabs, Mistral
Embeddings	OpenAI, Gemini, Azure, Cohere, Mistral, Jina, VoyageAI
Reranking	Cohere, Jina
Files	OpenAI, Anthropic, Gemini

The `Laravel\Ai\Enums\Lab` enum may be used to reference providers throughout your code instead of using plain strings:

```
1  use Laravel\Ai\Enums\Lab;  
2  
3  Lab::Anthropic;  
4  Lab::OpenAI;  
5  Lab::Gemini;  
6  // ...
```

Agents

Agents are the fundamental building block for interacting with AI providers in the Laravel AI SDK. Each agent is a dedicated PHP class that encapsulates the instructions, conversation context, tools, and output schema needed to interact with a large language model. Think of an agent as a specialized assistant — a sales coach, a document analyzer, a support bot — that you configure once and prompt as needed throughout your application.

You can create an agent via the `make:agent` Artisan command:

```
1  php artisan make:agent SalesCoach
2
3  php artisan make:agent SalesCoach --structured
```

Within the generated agent class, you can define the system prompt / instructions, message context, available tools, and output schema (if applicable):

```
1  <?php
2
3  namespace App\Ai\Agents;
4
5  use App\Ai\Tools\RetrievePreviousTranscripts;
6  use App\Models\History;
7  use App\Models\User;
8  use Illuminate\Contracts\JsonSchema\JsonSchema;
9  use Laravel\Ai\Contracts\Agent;
10 use Laravel\Ai\Contracts\Conversational;
11 use Laravel\Ai\Contracts\HasStructuredOutput;
12 use Laravel\Ai\Contracts\HasTools;
13 use Laravel\Ai\Messages\Message;
14 use Laravel\Ai\Promptable;
15 use Stringable;
16
17 class SalesCoach implements Agent, Conversational, HasTools, HasStructuredOutput
18 {
19     use Promptable;
20
21     public function __construct(public User $user) {}
22
23     /**
24      * Get the instructions that the agent should follow.
25      */
26     public function instructions(): Stringable|string
27     {
28         return 'You are a sales coach, analyzing transcripts and providing f
29     }
30
31     /**
32      * Get the list of messages comprising the conversation so far.
33      */
34     public function messages(): iterable
35     {
36         return History::where('user_id', $this->user->id)
37             ->latest()
38             ->limit(50)
39             ->get()
```

```
40         →reverse()
41         →map(function ($message) {
42             return new Message($message→role, $message→content);
43         })→all();
44     }
45
46     /**
47      * Get the tools available to the agent.
48      *
49      * @return Tool[]
50      */
51     public function tools(): iterable
```

Laravel

v12.x ▾



```
55         ];
56     }
57
58     /**
59      * Get the agent's structured output schema definition.
60      */
61     public function schema(JsonSchema $schema): array
62     {
63         return [
64             'feedback' => $schema→string()→required(),
65             'score' => $schema→integer()→min(1)→max(10)→required(),
66         ];
67     }
68 }
```

Prompting

To prompt an agent, first create an instance using the `make` method or standard instantiation, then call `prompt`:

```
1  $response = (new SalesCoach)
2      →prompt('Analyze this sales transcript... ');
3
4  $response = SalesCoach::make()
5      →prompt('Analyze this sales transcript... ');
6
7  return (string) $response;
```

The `make` method resolves your agent from the container, allowing automatic dependency injection. You may also pass arguments to the agent's constructor:

```
1 $agent = SalesCoach::make(user: $user);
```

By passing additional arguments to the `prompt` method, you may override the default provider, model, or HTTP timeout when prompting:

```
1 $response = (new SalesCoach)→prompt(  
2     'Analyze this sales transcript...',  
3     provider: Lab::Anthropic,  
4     model: 'claude-haiku-4-5-20251001',  
5     timeout: 120,  
6 );
```

Conversation Context

If your agent implements the `Conversational` interface, you may use the `messages` method to return the previous conversation context, if applicable:

```
1 use App\Models\History;  
2 use Laravel\Ai\Messages\Message;  
3  
4 /**  
5 * Get the list of messages comprising the conversation so far.  
6 */  
7 public function messages(): iterable  
8 {  
9     return History::where('user_id', $this→user→id)  
10        →latest()  
11        →limit(50)  
12        →get()  
13        →reverse()  
14        →map(function ($message) {  
15            return new Message($message→role, $message→content);  
16        })→all();  
17 }
```

Remembering Conversations

Before using the `RemembersConversations` trait, you should publish and run the AI SDK migrations using the `vendor:publish` Artisan command. These migrations will create the necessary database tables to store conversations.

If you would like Laravel to automatically store and retrieve conversation history for your agent, you may use the `RemembersConversations` trait. This trait provides a simple way to persist conversation messages to the database without manually implementing the `Conversational` interface:

```
1 <?php
2
3 namespace App\Ai\Agents;
4
5 use Laravel\Ai\Concerns\RemembersConversations;
6 use Laravel\Ai\Contracts\Agent;
7 use Laravel\Ai\Contracts\Conversational;
8 use Laravel\Ai\Promptable;
9
10 class SalesCoach implements Agent, Conversational
11 {
12     use Promptable, RemembersConversations;
13
14     /**
15      * Get the instructions that the agent should follow.
16      */
17     public function instructions(): string
18     {
19         return 'You are a sales coach...';
20     }
21 }
```

To start a new conversation for a user, call the `forUser` method before prompting:

```
1 $response = (new SalesCoach)→forUser($user)→prompt('Hello!');
2
3 $conversationId = $response→conversationId;
```

The conversation ID is returned on the response and can be stored for future reference, or you can retrieve all of a user's conversations from the `agent_conversations` table directly.

To continue an existing conversation, use the `continue` method:

```
1 $response = (new SalesCoach)
2     →continue($conversationId, as: $user)
3     →prompt('Tell me more about that.');
```

When using the `RemembersConversations` trait, previous messages are automatically loaded and included in the conversation context when prompting. New messages (both user and assistant) are automatically stored after each interaction.

Structured Output

If you would like your agent to return structured output, implement the `HasStructuredOutput` interface, which requires that your agent define a `schema` method:

```
1  <?php
2
3  namespace App\Ai\Agents;
4
5  use Illuminate\Contracts\JsonSchema\JsonSchema;
6  use Laravel\Ai\Contracts\Agent;
7  use Laravel\Ai\Contracts\HasStructuredOutput;
8  use Laravel\Ai\Promptable;
9
10 class SalesCoach implements Agent, HasStructuredOutput
11 {
12     use Promptable;
13
14     // ...
15
16     /**
17      * Get the agent's structured output schema definition.
18      */
19     public function schema(JsonSchema $schema): array
20     {
21         return [
22             'score' => $schema->integer()->required(),
23         ];
24     }
25 }
```

When prompting an agent that returns structured output, you can access the returned `StructuredAgentResponse` like an array:

```
1  $response = (new SalesCoach)->prompt('Analyze this sales transcript... ');
2
3  return $response['score'];
```

Attachments

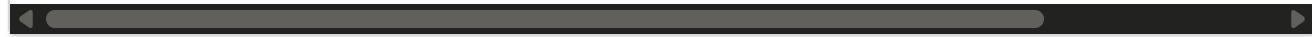
When prompting, you may also pass attachments with the prompt to allow the model to inspect images and documents:

```
1  use App\Ai\Agents\SalesCoach;
2  use Laravel\Ai\Files;
3
4  $response = (new SalesCoach)→prompt(
5      'Analyze the attached sales transcript... ',
6      attachments: [
7          Files\Document::fromStorage('transcript.pdf') // Attach a document f
8          Files\Document::fromPath('/home/laravel/transcript.md') // Attach a
9          $request→file('transcript'), // Attach an uploaded file...
10     ]
11 );
```



Likewise, the `Laravel\Ai\Files\Image` class may be used to attach images to a prompt:

```
1  use App\Ai\Agents\ImageAnalyzer;
2  use Laravel\Ai\Files;
3
4  $response = (new ImageAnalyzer)→prompt(
5      'What is in this image?',
6      attachments: [
7          Files\Image::fromStorage('photo.jpg') // Attach an image from a file
8          Files\Image::fromPath('/home/laravel/photo.jpg') // Attach an image
9          $request→file('photo'), // Attach an uploaded file...
10     ]
11 );
```



Streaming

You may stream an agent's response by invoking the `stream` method. The returned `StreamableAgentResponse` may be returned from a route to automatically send a streaming response (SSE) to the client:

```
1  use App\Ai\Agents\SalesCoach;
2
3  Route::get('/coach', function () {
4      return (new SalesCoach)→stream('Analyze this sales transcript... ');
5  });
```

The `then` method may be used to provide a closure that will be invoked when the entire response has been streamed to the client:

```
1  use App\Ai\Agents\SalesCoach;
2  use Laravel\Ai\Responses\StreamedAgentResponse;
3
4  Route::get('/coach', function () {
5      return (new SalesCoach)
6          ->stream('Analyze this sales transcript... ')
7          ->then(function (StreamedAgentResponse $response) {
8              // $response->text, $response->events, $response->usage ...
9          });
10});
```

Alternatively, you may iterate through the streamed events manually:

```
1  $stream = (new SalesCoach)->stream('Analyze this sales transcript... ');
2
3  foreach ($stream as $event) {
4      // ...
5 }
```

Streaming Using the Vercel AI SDK Protocol

You may stream the events using the [Vercel AI SDK stream protocol](#) by invoking the `usingVercelDataProtocol` method on the streamable response:

```
1  use App\Ai\Agents\SalesCoach;
2
3  Route::get('/coach', function () {
4      return (new SalesCoach)
5          ->stream('Analyze this sales transcript... ')
6          ->usingVercelDataProtocol();
7 });
```

Broadcasting

You may broadcast streamed events in a few different ways. First, you can simply invoke the `broadcast` or `broadcastNow` method on a streamed event:

```
1  use App\Ai\Agents\SalesCoach;
2  use Illuminate\Broadcasting\Channel;
```

```
3
4     $stream = (new SalesCoach)→stream('Analyze this sales transcript...');
5
6     foreach ($stream as $event) {
7         $event→broadcast(new Channel('channel-name'));
8     }
```

Or, you can invoke an agent's `broadcastOnQueue` method to queue the agent operation and broadcast the streamed events as they are available:

```
1     (new SalesCoach)→broadcastOnQueue(
2         'Analyze this sales transcript... '
3         new Channel('channel-name'),
4     );
```

Queueing

Using an agent's `queue` method, you may prompt the agent, but allow it to process the response in the background, keeping your application feeling fast and responsive. The `then` and `catch` methods may be used to register closures that will be invoked when a response is available or if an exception occurs:

```
1     use Illuminate\Http\Request;
2     use Laravel\Ai\Responses\AgentResponse;
3     use Throwable;
4
5     Route::post('/coach', function (Request $request) {
6         return (new SalesCoach)
7             →queue($request→input('transcript'))
8             →then(function (AgentResponse $response) {
9                 // ...
10            })
11            →catch(function (Throwable $e) {
12                // ...
13            });
14
15         return back();
16     });
```

Tools

Tools may be used to give agents additional functionality that they can utilize while responding to prompts. Tools can be created using the `make:tool` Artisan command:

```
1  php artisan make:tool RandomNumberGenerator
```

The generated tool will be placed in your application's `app/Ai/Tools` directory. Each tool contains a `handle` method that will be invoked by the agent when it needs to utilize the tool:

```
1  <?php
2
3  namespace App\Ai\Tools;
4
5  use Illuminate\Contracts\JsonSchema\JsonSchema;
6  use Laravel\Ai\Contracts\Tool;
7  use Laravel\Ai\Tools\Request;
8  use Stringable;
9
10 class RandomNumberGenerator implements Tool
11 {
12     /**
13      * Get the description of the tool's purpose.
14     */
15     public function description(): Stringable|string
16     {
17         return 'This tool may be used to generate cryptographically secure r
18     }
19
20     /**
21      * Execute the tool.
22     */
23     public function handle(Request $request): Stringable|string
24     {
25         return (string) random_int($request['min'], $request['max']);
26     }
27
28     /**
29      * Get the tool's schema definition.
30     */
31     public function schema(JsonSchema $schema): array
32     {
33         return [
34             'min' => $schema->integer()->min(0)->required(),
35             'max' => $schema->integer()->required(),
36         ];
37     }
38 }
```



Once you have defined your tool, you may return it from the `tools` method of any of your agents:

```
1  use App\Ai\Tools\RandomNumberGenerator;
2
3  /**
4   * Get the tools available to the agent.
5   *
6   * @return Tool[]
7   */
8  public function tools(): iterable
9  {
10     return [
11         new RandomNumberGenerator,
12     ];
13 }
```

Similarity Search

The `SimilaritySearch` tool allows agents to search for documents similar to a given query using vector embeddings stored in your database. This is useful for retrieval-augmented generation (RAG) when you want to give agents access to search your application's data.

The simplest way to create a similarity search tool is using the `usingModel` method with an Eloquent model that has vector embeddings:

```
1  use App\Models\Document;
2  use Laravel\Ai\Tools\SimilaritySearch;
3
4  public function tools(): iterable
5  {
6      return [
7          SimilaritySearch::usingModel(Document::class, 'embedding'),
8      ];
9  }
```

The first argument is the Eloquent model class, and the second argument is the column containing the vector embeddings.

You may also provide a minimum similarity threshold between `0.0` and `1.0` and a closure to customize the query:

```
1  SimilaritySearch::usingModel(
2      model: Document::class,
```

```
3     column: 'embedding',
4     minSimilarity: 0.7,
5     limit: 10,
6     query: fn ($query) => $query→where('published', true),
7   ),
```

For more control, you may create a similarity search tool with a custom closure that returns the search results:

```
1  use App\Models\Document;
2  use Laravel\AI\Tools\SimilaritySearch;
3
4  public function tools(): iterable
5  {
6    return [
7      new SimilaritySearch(using: function (string $query) {
8        return Document::query()
9          →where('user_id', $this→user→id)
10         →whereVectorSimilarTo('embedding', $query)
11         →limit(10)
12         →get();
13      }),
14    ];
15 }
```

You may customize the tool's description using the `withDescription` method:

```
1  SimilaritySearch::usingModel(Document::class, 'embedding')
2    →withDescription('Search the knowledge base for relevant articles.'),
```

Provider Tools

Provider tools are special tools implemented natively by AI providers, offering capabilities like web searching, URL fetching, and file searching. Unlike regular tools, provider tools are executed by the provider itself rather than your application.

Provider tools can be returned by your agent's `tools` method.

Web Search

The `WebSearch` provider tool allows agents to search the web for real-time information. This is useful for answering questions about current events, recent data, or topics that may have changed since the model's training cutoff.

Supported Providers: Anthropic, OpenAI, Gemini

```
1  use Laravel\Ai\Providers\Tools\WebSearch;
2
3  public function tools(): iterable
4  {
5      return [
6          new WebSearch,
7      ];
8  }
```

You may configure the web search tool to limit the number of searches or restrict results to specific domains:

```
1  (new WebSearch)→max(5)→allow(['laravel.com', 'php.net']),
```

To refine search results based on user location, use the `location` method:

```
1  (new WebSearch)→location(
2      city: 'New York',
3      region: 'NY',
4      country: 'US'
5  );
```

Web Fetch

The `WebFetch` provider tool allows agents to fetch and read the contents of web pages. This is useful when you need the agent to analyze specific URLs or retrieve detailed information from known web pages.

Supported providers: Anthropic, Gemini

```
1  use Laravel\Ai\Providers\Tools\WebFetch;
2
3  public function tools(): iterable
4  {
5      return [
6          new WebFetch,
7      ];
8  }
```

You may configure the web fetch tool to limit the number of fetches or restrict to specific domains:

```
1 (new WebFetch)→max(3)→allow(['docs.laravel.com']),
```

File Search

The `FileSearch` provider tool allows agents to search through [files](#) stored in [vector stores](#). This enables retrieval-augmented generation (RAG) by allowing the agent to search your uploaded documents for relevant information.

Supported providers: OpenAI, Gemini

```
1 use Laravel\Ai\Providers\Tools\FileSearch;
2
3 public function tools(): iterable
4 {
5     return [
6         new FileSearch(stores: ['store_id']),
7     ];
8 }
```

You may provide multiple vector store IDs to search across multiple stores:

```
1 new FileSearch(stores: ['store_1', 'store_2']);
```

If your files have [metadata](#), you may filter the search results by providing a `where` argument. For simple equality filters, pass an array:

```
1 new FileSearch(stores: ['store_id'], where: [
2     'author' => 'Taylor Otwell',
3     'year' => 2026,
4 ]);
```

For more complex filters, you may pass a closure that receives a `FileSearchQuery` instance:

```
1 use Laravel\Ai\Providers\Tools\FileSearchQuery;
2
3 new FileSearch(stores: ['store_id'], where: fn (FileSearchQuery $query) =>
4     $query→where('author', 'Taylor Otwell')
```

```
5      →whereNot('status', 'draft')
6      →whereIn('category', ['news', 'updates'])
7  );
```

Middleware

Agents support middleware, allowing you to intercept and modify prompts before they are sent to the provider. To add middleware to an agent, implement the `HasMiddleware` interface and define a `middleware` method that returns an array of middleware classes:

```
1  <?php
2
3  namespace App\Ai\Agents;
4
5  use Laravel\Ai\Contracts\Agent;
6  use Laravel\Ai\Contracts\HasMiddleware;
7  use Laravel\Ai\Promptable;
8
9  class SalesCoach implements Agent, HasMiddleware
10 {
11     use Promptable;
12
13     // ...
14
15     /**
16      * Get the agent's middleware.
17      */
18     public function middleware(): array
19     {
20         return [
21             new LogPrompts,
22         ];
23     }
24 }
```

Each middleware class should define a `handle` method that receives the `AgentPrompt` and a `Closure` to pass the prompt to the next middleware:

```
1  <?php
2
3  namespace App\Ai\Middleware;
4
5  use Closure;
6  use Laravel\Ai\Prompts\AgentPrompt;
7
8  class LogPrompts
```

```
9     {
10         /**
11          * Handle the incoming prompt.
12         */
13         public function handle(AgentPrompt $prompt, Closure $next)
14         {
15             Log::info('Prompting agent', ['prompt' => $prompt->prompt]);
16
17             return $next($prompt);
18         }
19     }
```

You may use the `then` method on the response to execute code after the agent has finished processing. This works for both synchronous and streaming responses:

```
1  public function handle(AgentPrompt $prompt, Closure $next)
2  {
3      return $next($prompt)->then(function (AgentResponse $response) {
4          Log::info('Agent responded', ['text' => $response->text]);
5      });
6  }
```

Anonymous Agents

Sometimes you may want to quickly interact with a model without creating a dedicated agent class. You can create an ad-hoc, anonymous agent using the `agent` function:

```
1  use function Laravel\Ai\{agent};
2
3  $response = agent(
4      instructions: 'You are an expert at software development.',
5      messages: [],
6      tools: [],
7  )->prompt('Tell me about Laravel')
```

Anonymous agents may also produce structured output:

```
1  use Illuminate\Contracts\JsonSchema\JsonSchema;
2
3  use function Laravel\Ai\{agent};
4
5  $response = agent(
6      schema: fn (JsonSchema $schema) => [
7          'number' => $schema->integer()->required(),
```

```
8      ],
9  )->prompt('Generate a random number less than 100')
```

Agent Configuration

You may configure text generation options for an agent using PHP attributes. The following attributes are available:

MaxSteps: The maximum number of steps the agent may take when using tools.

MaxTokens: The maximum number of tokens the model may generate.

Model: The model the agent should use.

Provider: The AI provider (or providers for failover) to use for the agent.

Temperature: The sampling temperature to use for generation (0.0 to 1.0).

Timeout: The HTTP timeout in seconds for agent requests (default: 60).

UseCheapestModel: Use the provider's cheapest text model for cost optimization.

UseSmartestModel: Use the provider's most capable text model for complex tasks.

```
1  <?php
2
3  namespace App\Ai\Agents;
4
5  use Laravel\Ai\Attributes\MaxSteps;
6  use Laravel\Ai\Attributes\MaxTokens;
7  use Laravel\Ai\Attributes\Model;
8  use Laravel\Ai\Attributes\Provider;
9  use Laravel\Ai\Attributes\Temperature;
10 use Laravel\Ai\Attributes\Timeout;
11 use Laravel\Ai\Contracts\Agent;
12 use Laravel\Ai\Enums\Lab;
13 use Laravel\Ai\Promptable;
14
15 #[Provider(Lab::Anthropic)]
16 #[Model('claude-haiku-4-5-20251001')]
17 #[MaxSteps(10)]
18 #[MaxTokens(4096)]
19 #[Temperature(0.7)]
20 #[Timeout(120)]
21 class SalesCoach implements Agent
22 {
23     use Promptable;
24
25     // ...
26 }
```

The `UseCheapestModel` and `UseSmartestModel` attributes allow you to automatically select the most cost-effective or most capable model for a given provider without specifying a model name. This is useful when you want to optimize for cost or capability across different providers:

```
1  use Laravel\Ai\Attributes\UseCheapestModel;
2  use Laravel\Ai\Attributes\UseSmartestModel;
3  use Laravel\Ai\Contracts\Agent;
4  use Laravel\Ai\Promptable;
5
6  #[UseCheapestModel]
7  class SimpleSummarizer implements Agent
8  {
9      use Promptable;
10
11     // Will use the cheapest model (e.g., Haiku) ...
12 }
13
14 #[UseSmartestModel]
15 class ComplexReasoner implements Agent
16 {
17     use Promptable;
18
19     // Will use the most capable model (e.g., Opus) ...
20 }
```

Images

The `Laravel\Ai\Image` class may be used to generate images using the `openai`, `gemini`, or `xai` providers:

```
1  use Laravel\Ai\Image;
2
3  $image = Image::of('A donut sitting on the kitchen counter')->generate();
4
5  $rawContent = (string) $image;
```

The `square`, `portrait`, and `landscape` methods may be used to control the aspect ratio of the image, while the `quality` method may be used to guide the model on final image quality (`high`, `medium`, `low`). The `timeout` method may be used to specify the HTTP timeout in seconds:

```
1  use Laravel\Ai\Image;
2
3  $image = Image::of('A donut sitting on the kitchen counter')
4      →quality('high')
5      →landscape()
6      →timeout(120)
7      →generate();
```

You may attach reference images using the `attachments` method:

```
1  use Laravel\Ai\Files;
2  use Laravel\Ai\Image;
3
4  $image = Image::of('Update this photo of me to be in the style of an impress
5      →attachments([
6          Files\Image::fromStorage('photo.jpg'),
7          // Files\Image::fromPath('/home/laravel/photo.jpg'),
8          // Files\Image::fromUrl('https://example.com/photo.jpg'),
9          // $request→file('photo'),
10     ])
11     →landscape()
12     →generate();
```

Generated images may be easily stored on the default disk configured in your application's `config/filesystems.php` configuration file:

```
1  $image = Image::of('A donut sitting on the kitchen counter');
2
3  $path = $image→store();
4  $path = $image→storeAs('image.jpg');
5  $path = $image→storePublicly();
6  $path = $image→storePubliclyAs('image.jpg');
```

Image generation may also be queued:

```
1  use Laravel\Ai\Image;
2  use Laravel\Ai\Responses\ImageResponse;
3
4  Image::of('A donut sitting on the kitchen counter')
5      →portrait()
6      →queue()
7      →then(function (ImageResponse $image) {
```

```
8     $path = $image→store();
9
10    // ...
11 };
```

Audio

The `Laravel\Ai\Audio` class may be used to generate audio from the given text:

```
1 use Laravel\Ai\Audio;
2
3 $audio = Audio::of('I love coding with Laravel.')->generate();
4
5 $rawContent = (string) $audio;
```

The `male`, `female`, and `voice` methods may be used to determine the voice of the generated audio:

```
1 $audio = Audio::of('I love coding with Laravel.')
2   →female()
3   →generate();
4
5 $audio = Audio::of('I love coding with Laravel.')
6   →voice('voice-id-or-name')
7   →generate();
```

Similarly, the `instructions` method may be used to dynamically coach the model on how the generated audio should sound:

```
1 $audio = Audio::of('I love coding with Laravel.')
2   →female()
3   →instructions('Said like a pirate')
4   →generate();
```

Generated audio may be easily stored on the default disk configured in your application's `config/filesystems.php` configuration file:

```
1 $audio = Audio::of('I love coding with Laravel.')->generate();
2
3 $path = $audio→store();
```

```
4     $path = $audio→storeAs('audio.mp3');
5     $path = $audio→storePublicly();
6     $path = $audio→storePubliclyAs('audio.mp3');
```

Audio generation may also be queued:

```
1  use Laravel\Ai\Audio;
2  use Laravel\Ai\Responses\AudioResponse;
3
4  Audio::of('I love coding with Laravel.')
5      →queue()
6      →then(function (AudioResponse $audio) {
7          $path = $audio→store();
8
9          // ...
10     });
});
```

Transcriptions

The `Laravel\Ai\Transcription` class may be used to generate a transcript of the given audio:

```
1  use Laravel\Ai\Transcription;
2
3  $transcript = Transcription::fromPath('/home/laravel/audio.mp3')→generate();
4  $transcript = Transcription::fromStorage('audio.mp3')→generate();
5  $transcript = Transcription::fromUpload($request→file('audio'))→generate();
6
7  return (string) $transcript;
```

The `diarize` method may be used to indicate you would like the response to include the diarized transcript in addition to the raw text transcript, allowing you to access the segmented transcript by speaker:

```
1  $transcript = Transcription::fromStorage('audio.mp3')
2      →diarize()
3      →generate();
```

Transcription generation may also be queued:

```
1  use Laravel\Ai\Transcription;
2  use Laravel\Ai\Responses\TranscriptionResponse;
3
4  Transcription::fromStorage('audio.mp3')
5      →queue()
6      →then(function (TranscriptionResponse $transcript) {
7          // ...
8      });

```

Embeddings

You may easily generate vector embeddings for any given string using the new `toEmbeddings` method available via Laravel's `Stringable` class:

```
1  use Illuminate\Support\Str;
2
3  $embeddings = Str::of('Napa Valley has great wine.')->toEmbeddings();
```

Alternatively, you may use the `Embeddings` class to generate embeddings for multiple inputs at once:

```
1  use Laravel\Ai\Embeddings;
2
3  $response = Embeddings::for([
4      'Napa Valley has great wine.',
5      'Laravel is a PHP framework.',
6  ])->generate();
7
8  $response->embeddings; // [[0.123, 0.456, ...], [0.789, 0.012, ...]]
```

You may specify the dimensions and provider for the embeddings:

```
1  $response = Embeddings::for(['Napa Valley has great wine.'])
2      →dimensions(1536)
3      →generate(Lab::OpenAI, 'text-embedding-3-small');
```

Querying Embeddings

Once you have generated embeddings, you will typically store them in a `vector` column in your database for later querying. Laravel provides native support for vector columns on

PostgreSQL via the `pgvector` extension. To get started, define a `vector` column in your migration, specifying the number of dimensions:

```
1 Schema::ensureVectorExtensionExists();
2
3 Schema::create('documents', function (Blueprint $table) {
4     $table->id();
5     $table->string('title');
6     $table->text('content');
7     $table->vector('embedding', dimensions: 1536);
8     $table->timestamps();
9 });
```

You may also add a vector index to speed up similarity searches. When calling `index` on a vector column, Laravel will automatically create an HNSW index with cosine distance:

```
1 $table->vector('embedding', dimensions: 1536)->index();
```

On your Eloquent model, you should cast the vector column to an `array`:

```
1 protected function casts(): array
2 {
3     return [
4         'embedding' => 'array',
5     ];
6 }
```

To query for similar records, use the `whereVectorSimilarTo` method. This method filters results by a minimum cosine similarity (between `0.0` and `1.0`, where `1.0` is identical) and orders the results by similarity:

```
1 use App\Models\Document;
2
3 $documents = Document::query()
4     ->whereVectorSimilarTo('embedding', $queryEmbedding, minSimilarity: 0.4)
5     ->limit(10)
6     ->get();
```

The `$queryEmbedding` may be an array of floats or a plain string. When a string is given, Laravel will automatically generate embeddings for it:

```
1 $documents = Document::query()  
2     →whereVectorSimilarTo('embedding', 'best wineries in Napa Valley')  
3     →limit(10)  
4     →get();
```

If you need more control, you may use the lower-level `whereVectorDistanceLessThan`, `selectVectorDistance`, and `orderByVectorDistance` methods independently:

```
1 $documents = Document::query()  
2     →select('*')  
3     →selectVectorDistance('embedding', $queryEmbedding, as: 'distance')  
4     →whereVectorDistanceLessThan('embedding', $queryEmbedding, maxDistance:  
5         →orderByVectorDistance('embedding', $queryEmbedding)  
6     →limit(10)  
7     →get();
```

If you would like to give an agent the ability to perform similarity searches as a tool, check out the [Similarity Search](#) tool documentation.

Vector queries are currently only supported on PostgreSQL connections using the `pgvector` extension.

Caching Embeddings

Embedding generation can be cached to avoid redundant API calls for identical inputs. To enable caching, set the `ai.caching.embeddings.cache` configuration option to `true`:

```
1 'caching' => [  
2     'embeddings' => [  
3         'cache' => true,  
4         'store' => env('CACHE_STORE', 'database'),  
5         // ...  
6     ],  
7 ];
```

When caching is enabled, embeddings are cached for 30 days. The cache key is based on the provider, model, dimensions, and input content, ensuring that identical requests return cached results while different configurations generate fresh embeddings.

You may also enable caching for a specific request using the `cache` method, even when global caching is disabled:

```
1 $response = Embeddings::for(['Napa Valley has great wine.'])
2     →cache()
3     →generate();
```

You may specify a custom cache duration in seconds:

```
1 $response = Embeddings::for(['Napa Valley has great wine.'])
2     →cache(seconds: 3600) // Cache for 1 hour
3     →generate();
```

The `toEmbeddings` Stringable method also accepts a `cache` argument:

```
1 // Cache with default duration...
2 $embeddings = Str::of('Napa Valley has great wine.')→toEmbeddings(cache: true)
3
4 // Cache for a specific duration...
5 $embeddings = Str::of('Napa Valley has great wine.')→toEmbeddings(cache: 3600)
```

Reranking

Reranking allows you to reorder a list of documents based on their relevance to a given query. This is useful for improving search results by using semantic understanding:

The `Laravel\Ai\Reranking` class may be used to rerank documents:

```
1 use Laravel\Ai\Reranking;
2
3 $response = Reranking::of([
4     'Django is a Python web framework.',
5     'Laravel is a PHP web application framework.',
6     'React is a JavaScript library for building user interfaces.',
7 ])→rerank('PHP frameworks');
8
9 // Access the top result...
10 $response→first()→document; // "Laravel is a PHP web application framework"
11 $response→first()→score;    // 0.95
12 $response→first()→index;   // 1 (original position)
```

The `limit` method may be used to restrict the number of results returned:

```
1 $response = Reranking::of($documents)
2     ->limit(5)
3     ->rerank('search query');
```

Reranking Collections

For convenience, Laravel collections may be reranked using the `rerank` macro. The first

~~argument specifies which field(s) to use for reranking, and the second argument is the query:~~

```
1 // Rerank by a single field...
2 $posts = Post::all()
3     ->rerank('body', 'Laravel tutorials');
4
5 // Rerank by multiple fields (sent as JSON)...
6 $reranked = $posts->rerank(['title', 'body'], 'Laravel tutorials');
7
8 // Rerank using a closure to build the document...
9 $reranked = $posts->rerank(
10     fn ($post) => $post->title.': '.$post->body,
11     'Laravel tutorials'
12 );
```

You may also limit the number of results and specify a provider:

```
1 $reranked = $posts->rerank(
2     by: 'content',
3     query: 'Laravel tutorials',
4     limit: 10,
5     provider: Lab::Cohere
6 );
```

Files

The `Laravel\AI\Files` class or the individual file classes may be used to store files with your AI provider for later use in conversations. This is useful for large documents or files you want to reference multiple times without re-uploading:

```
1  use Laravel\Ai\Files\Document;
2  use Laravel\Ai\Files\Image;
3
4  // Store a file from a local path...
5  $response = Document::fromPath('/home/laravel/document.pdf')→put();
6  $response = Image::fromPath('/home/laravel/photo.jpg')→put();
7
8  // Store a file that is stored on a filesystem disk...
9  $response = Document::fromStorage('document.pdf', disk: 'local')→put();
10 $response = Image::fromStorage('photo.jpg', disk: 'local')→put();
11
12 // Store a file that is stored on a remote URL...
13 $response = Document::fromUrl('https://example.com/document.pdf')→put();
14 $response = Image::fromUrl('https://example.com/photo.jpg')→put();
15
16 return $response→id;
```

You may also store raw content or uploaded files:

```
1  use Laravel\Ai\Files;
2  use Laravel\Ai\Files\Document;
3
4  // Store raw content...
5  $stored = Document::fromString('Hello, World!', 'text/plain')→put();
6
7  // Store an uploaded file...
8  $stored = Document::fromUpload($request→file('document'))→put();
```

Once a file has been stored, you may reference the file when generating text via agents instead of re-uploading the file:

```
1  use App\Ai\Agents\SalesCoach;
2  use Laravel\Ai\Files;
3
4  $response = (new SalesCoach)→prompt(
5      'Analyze the attached sales transcript...'
6      attachments: [
7          Files\Document::fromId('file-id') // Attach a stored document...
8      ]
9  );
```

To retrieve a previously stored file, use the `get` method on a file instance:

```
1  use Laravel\Ai\Files\Document;
2
3  $file = Document::fromId('file-id')→get();
4
5  $file→id;
6  $file→mimeType();
```

To delete a file from the provider, use the `delete` method:

```
1  Document::fromId('file-id')→delete();
```

By default, the `Files` class uses the default AI provider configured in your application's `config/ai.php` configuration file. For most operations, you may specify a different provider using the `provider` argument:

```
1  $response = Document::fromPath(
2      '/home/laravel/document.pdf'
3  )→put(provider: Lab::Anthropic);
```

Using Stored Files in Conversations

Once a file has been stored with a provider, you may reference it in agent conversations using the `fromId` method on the `Document` or `Image` classes:

```
1  use App\Ai\Agents\DocumentAnalyzer;
2  use Laravel\Ai\Files;
3  use Laravel\Ai\Files\Document;
4
5  $stored = Document::fromPath('/path/to/report.pdf')→put();
6
7  $response = (new DocumentAnalyzer)→prompt(
8      'Summarize this document.',
9      attachments: [
10          Document::fromId($stored→id),
11      ],
12  );
```

Similarly, stored images may be referenced using the `Image` class:

```
1  use Laravel\Ai\Files;
```

```
2  use Laravel\Ai\Files\Image;
3
4  $stored = Image::fromPath('/path/to/photo.jpg')->put();
5
6  $response = (new ImageAnalyzer)->prompt(
7      'What is in this image?',
8      attachments: [
9          Image::fromId($stored->id),
10     ],
11 );
```

Vector Stores

Vector stores allow you to create searchable collections of files that can be used for retrieval-augmented generation (RAG). The `Laravel\Ai\Stores` class provides methods for creating, retrieving, and deleting vector stores:

```
1  use Laravel\Ai\Stores;
2
3  // Create a new vector store...
4  $store = Stores::create('Knowledge Base');
5
6  // Create a store with additional options...
7  $store = Stores::create(
8      name: 'Knowledge Base',
9      description: 'Documentation and reference materials.',
10     expiresWhenIdleFor: days(30),
11 );
12
13 return $store->id;
```

To retrieve an existing vector store by its ID, use the `get` method:

```
1  use Laravel\Ai\Stores;
2
3  $store = Stores::get('store_id');
4
5  $store->id;
6  $store->name;
7  $store->fileCounts;
8  $store->ready;
```

To delete a vector store, use the `delete` method on the `Stores` class or the store instance:

```
1  use Laravel\Ai\Stores;  
2  
3  // Delete by ID...  
4  Stores::delete('store_id');  
5  
6  // Or delete via a store instance...  
7  $store = Stores::get('store_id');  
8  
9  $store->delete();
```

Adding Files to Stores

Once you have a vector store, you may add [files](#) to it using the `add` method. Files added to a store are automatically indexed for semantic searching using the [file search provider tool](#):

```
1  use Laravel\Ai\Files\Document;  
2  use Laravel\Ai\Stores;  
3  
4  $store = Stores::get('store_id');  
5  
6  // Add a file that has already been stored with the provider...  
7  $document = $store->add('file_id');  
8  $document = $store->add(Document::fromId('file_id'));  
9  
10 // Or, store and add a file in one step...  
11 $document = $store->add(Document::fromPath('/path/to/document.pdf'));  
12 $document = $store->add(Document::fromStorage('manual.pdf'));  
13 $document = $store->add($request->file('document'));  
14  
15 $document->id;  
16 $document->fileId;
```

Typically, when adding previously stored files to vector stores, the returned document ID will match the file's previously assigned ID; however, some vector storage providers may return a new, different "document ID". Therefore, it's recommended that you always store both IDs in your database for future reference.

You may attach metadata to files when adding them to a store. This metadata can later be used to filter search results when using the [file search provider tool](#):

```
1 $store→add(Document::fromPath('/path/to/document.pdf'), metadata: [
2     'author' => 'Taylor Otwell',
3     'department' => 'Engineering',
4     'year' => 2026,
5 ]);
```

To remove a file from a store, use the `remove` method:

```
1 $store→remove('file_id');
```

Removing a file from a vector store does not remove it from the provider's [file storage](#). To remove a file from the vector store and delete it permanently from file storage, use the `deleteFile` argument:

```
1 $store→remove('file_abci23', deleteFile: true);
```

Failover

When prompting or generating other media, you may provide an array of providers / models to automatically failover to a backup provider / model if a service interruption or rate limit is encountered on the primary provider:

```
1 use App\Ai\Agents\SalesCoach;
2 use Laravel\Ai\Image;
3
4 $response = (new SalesCoach)→prompt(
5     'Analyze this sales transcript... ',
6     provider: [Lab::OpenAI, Lab::Anthropic],
7 );
8
9 $image = Image::of('A donut sitting on the kitchen counter')
10 →generate(provider: [Lab::Gemini, Lab::xAI]);
```

Testing

Agents

To fake an agent's responses during tests, call the `fake` method on the agent class. You may optionally provide an array of responses or a closure:

```
1  use App\Ai\Agents\SalesCoach;
2  use Laravel\Ai\Prompts\AgentPrompt;
3
4  // Automatically generate a fixed response for every prompt...
5  SalesCoach::fake();
6
7  // Provide a list of prompt responses...
8  SalesCoach::fake([
9      'First response',
10     'Second response',
11 ]);
12
13 // Dynamically handle prompt responses based on the incoming prompt...
14 SalesCoach::fake(function (AgentPrompt $prompt) {
15     return 'Response for: '.$prompt→prompt;
16 });
```

When `Agent::fake()` is invoked on an agent that returns structured output, Laravel will automatically generate fake data that matches your agent's defined output schema.

After prompting the agent, you may make assertions about the prompts that were received:

```
1  use Laravel\Ai\Prompts\AgentPrompt;
2
3  SalesCoach::assertPrompted('Analyze this ...');
4
5  SalesCoach::assertPrompted(function (AgentPrompt $prompt) {
6      return $prompt→contains('Analyze');
7  });
8
9  SalesCoach::assertNotPrompted('Missing prompt');
10
11 SalesCoach::assertNeverPrompted();
```

For queued agent invocations, use the queued assertion methods:

```
1  use Laravel\Ai\QueueAgentPrompt;
```

```
2
3     SalesCoach::assertQueued('Analyze this ...');
4
5     SalesCoach::assertQueued(function (QueuedAgentPrompt $prompt) {
6         return $prompt->contains('Analyze');
7     });
8
9     SalesCoach::assertNotQueued('Missing prompt');
10
11    SalesCoach::assertNeverQueued();
```

To ensure all agent invocations have a corresponding fake response, you may use `preventStrayPrompts`. If an agent is invoked without a defined fake response, an exception will be thrown:

```
1     SalesCoach::fake()->preventStrayPrompts();
```

Images

Image generations may be faked by invoking the `fake` method on the `Image` class. Once image has been faked, various assertions may be performed against the recorded image generation prompts:

```
1     use Laravel\Ai\Image;
2     use Laravel\Ai\Prompts\ImagePrompt;
3     use Laravel\Ai\Prompts\QueuedImagePrompt;
4
5     // Automatically generate a fixed response for every prompt...
6     Image::fake();
7
8     // Provide a list of prompt responses...
9     Image::fake([
10         base64_encode($firstImage),
11         base64_encode($secondImage),
12     ]);
13
14     // Dynamically handle prompt responses based on the incoming prompt...
15     Image::fake(function (ImagePrompt $prompt) {
16         return base64_encode('... ');
17     });

```

After generating images, you may make assertions about the prompts that were received:

```
1 Image::assertGenerated(function (ImagePrompt $prompt) {
2     return $prompt->contains('sunset') && $prompt->isLandscape();
3 });
4
5 Image::assertNotGenerated('Missing prompt');
6
7 Image::assertNothingGenerated();
```

For queued image generations, use the queued assertion methods:

```
1 Image::assertQueued(
2     fn (QueuedImagePrompt $prompt) => $prompt->contains('sunset')
3 );
4
5 Image::assertNotQueued('Missing prompt');
6
7 Image::assertNothingQueued();
```

To ensure all image generations have a corresponding fake response, you may use `preventStrayImages`. If an image is generated without a defined fake response, an exception will be thrown:

```
1 Image::fake()->preventStrayImages();
```

Audio

Audio generations may be faked by invoking the `fake` method on the `Audio` class. Once audio has been faked, various assertions may be performed against the recorded audio generation prompts:

```
1 use Laravel\Ai\Audio;
2 use Laravel\Ai\Prompts\AudioPrompt;
3 use Laravel\Ai\Prompts\QueuedAudioPrompt;
4
5 // Automatically generate a fixed response for every prompt...
6 Audio::fake();
7
8 // Provide a list of prompt responses...
9 Audio::fake([
10     base64_encode($firstAudio),
11     base64_encode($secondAudio),
12 ]);
```

```
13
14 // Dynamically handle prompt responses based on the incoming prompt...
15 Audio::fake(function (AudioPrompt $prompt) {
16     return base64_encode('...');
17 });
```

After generating audio, you may make assertions about the prompts that were received:

```
1 Audio::assertGenerated(function (AudioPrompt $prompt) {
2     return $prompt→contains('Hello') && $prompt→isFemale();
3 });
4
5 Audio::assertNotGenerated('Missing prompt');
6
7 Audio::assertNothingGenerated();
```

For queued audio generations, use the queued assertion methods:

```
1 Audio::assertQueued(
2     fn (QueuedAudioPrompt $prompt) => $prompt→contains('Hello')
3 );
4
5 Audio::assertNotQueued('Missing prompt');
6
7 Audio::assertNothingQueued();
```

To ensure all audio generations have a corresponding fake response, you may use `preventStrayAudio`. If audio is generated without a defined fake response, an exception will be thrown:

```
1 Audio::fake()→preventStrayAudio();
```

Transcriptions

Transcription generations may be faked by invoking the `fake` method on the `Transcription` class. Once transcription has been faked, various assertions may be performed against the recorded transcription generation prompts:

```
1 use Laravel\Ai\Transcription;
2 use Laravel\Ai\Prompts\TranscriptionPrompt;
3 use Laravel\Ai\Prompts\QueuedTranscriptionPrompt;
```

```
4
5 // Automatically generate a fixed response for every prompt...
6 Transcription::fake();
7
8 // Provide a list of prompt responses...
9 Transcription::fake([
10     'First transcription text.',
11     'Second transcription text.',
12 ]);
13
14 // Dynamically handle prompt responses based on the incoming prompt...
15 Transcription::fake(function (TranscriptionPrompt $prompt) {
16     return 'Transcribed text...';
17});
```

After generating transcriptions, you may make assertions about the prompts that were received:

```
1 Transcription::assertGenerated(function (TranscriptionPrompt $prompt) {
2     return $prompt→language === 'en' && $prompt→isDiarized();
3 });
4
5 Transcription::assertNotGenerated(
6     fn (TranscriptionPrompt $prompt) => $prompt→language === 'fr'
7 );
8
9 Transcription::assertNothingGenerated();
```

For queued transcription generations, use the queued assertion methods:

```
1 Transcription::assertQueued(
2     fn (QueuedTranscriptionPrompt $prompt) => $prompt→isDiarized()
3 );
4
5 Transcription::assertNotQueued(
6     fn (QueuedTranscriptionPrompt $prompt) => $prompt→language === 'fr'
7 );
8
9 Transcription::assertNothingQueued();
```

To ensure all transcription generations have a corresponding fake response, you may use `preventStrayTranscriptions`. If a transcription is generated without a defined fake response, an exception will be thrown:

```
1 Transcription::fake()→preventStrayTranscriptions();
```

Embeddings

Embeddings generations may be faked by invoking the `fake` method on the `Embeddings` class. Once embeddings has been faked, various assertions may be performed against the recorded embeddings generation prompts:

```
1 use Laravel\Ai\Embeddings;
2 use Laravel\Ai\Prompts\EmbeddingsPrompt;
3 use Laravel\Ai\Prompts\QueueEmbeddingsPrompt;
4
5 // Automatically generate fake embeddings of the proper dimensions for every
6 Embeddings::fake();
7
8 // Provide a list of prompt responses ...
9 Embeddings::fake([
10     [$firstEmbeddingVector],
11     [$secondEmbeddingVector],
12 ]);
13
14 // Dynamically handle prompt responses based on the incoming prompt ...
15 Embeddings::fake(function (EmbeddingsPrompt $prompt) {
16     return array_map(
17         fn () => Embeddings::fakeEmbedding($prompt→dimensions),
18         $prompt→inputs
19     );
20});
```



After generating embeddings, you may make assertions about the prompts that were received:

```
1 Embeddings::assertGenerated(function (EmbeddingsPrompt $prompt) {
2     return $prompt→contains('Laravel') && $prompt→dimensions === 1536;
3 });
4
5 Embeddings::assertNotGenerated(
6     fn (EmbeddingsPrompt $prompt) => $prompt→contains('Other')
7 );
8
9 Embeddings::assertNothingGenerated();
```

For queued embeddings generations, use the queued assertion methods:

```
1 Embeddings::assertQueued(
2     fn (QueuedEmbeddingsPrompt $prompt) => $prompt->contains('Laravel')
3 );
4
5 Embeddings::assertNotQueued(
6     fn (QueuedEmbeddingsPrompt $prompt) => $prompt->contains('Other')
7 );
8
9 Embeddings::assertNothingQueued();
```

To ensure all embeddings generations have a corresponding fake response, you may use `preventStrayEmbeddings`. If embeddings are generated without a defined fake response, an exception will be thrown:

```
1 Embeddings::fake()->preventStrayEmbeddings();
```

Reranking

Reranking operations may be faked by invoking the `fake` method on the `Reranking` class:

```
1 use Laravel\Ai\Reranking;
2 use Laravel\Ai\Prompts\RerankingPrompt;
3 use Laravel\Ai\Responses\Data\RankedDocument;
4
5 // Automatically generate a fake reranked responses ...
6 Reranking::fake();
7
8 // Provide custom responses ...
9 Reranking::fake([
10     [
11         new RankedDocument(index: 0, document: 'First', score: 0.95),
12         new RankedDocument(index: 1, document: 'Second', score: 0.80),
13     ],
14 ]);
```

After reranking, you may make assertions about the operations that were performed:

```
1 Reranking::assertReranked(function (RerankingPrompt $prompt) {
2     return $prompt->contains('Laravel') && $prompt->limit === 5;
3 });
4
5 Reranking::assertNotReranked()
```

```
6     fn (RerankingPrompt $prompt) => $prompt->contains('Django')
7 );
8
9 Reranking::assertNothingReranked();
```

Files

File operations may be faked by invoking the `fake` method on the `Files` class:

```
1 use Laravel\Ai\Files;
2
3 Files::fake();
```

Once file operations have been faked, you may make assertions about the uploads and deletions that occurred:

```
1 use Laravel\Ai\Contracts\Files\StorableFile;
2 use Laravel\Ai\Files\Document;
3
4 // Store files...
5 Document::fromString('Hello, Laravel!', mime: 'text/plain')
6     ->as('hello.txt')
7     ->put();
8
9 // Make assertions...
10 Files::assertStored(fn (StorableFile $file) =>
11     (string) $file === 'Hello, Laravel!' &&
12     $file->mimeType() === 'text/plain';
13 );
14
15 Files::assertNotStored(fn (StorableFile $file) =>
16     (string) $file === 'Hello, World!'
17 );
18
19 Files::assertNothingStored();
```

For asserting against file deletions, you may pass a file ID:

```
1 Files::assertDeleted('file-id');
2 Files::assertNotDeleted('file-id');
3 Files::assertNothingDeleted();
```

Vector Stores

Vector store operations may be faked by invoking the `fake` method on the `Stores` class.

Faking stores will also fake [file operations](#) automatically:

```
1  use Laravel\Ai\Stores;  
2  
3  Stores::fake();
```

Once store operations have been faked, you may make assertions about the stores that were created or deleted:

```
1  use Laravel\Ai\Stores;  
2  
3  // Create store...  
4  $store = Stores::create('Knowledge Base');  
5  
6  // Make assertions...  
7  Stores::assertCreated('Knowledge Base');  
8  
9  Stores::assertCreated(fn (string $name, ?string $description) =>  
10    $name === 'Knowledge Base'  
11  );  
12  
13  Stores::assertNotCreated('Other Store');  
14  
15  Stores::assertNothingCreated();
```

For asserting against store deletions, you may provide the store ID:

```
1  Stores::assertDeleted('store_id');  
2  Stores::assertNotDeleted('other_store_id');  
3  Stores::assertNothingDeleted();
```

To assert files were added or removed from a store, use the assertion methods on a given `Store` instance:

```
1  Stores::fake();  
2  
3  $store = Stores::get('store_id');  
4  
5  // Add / remove files...  
6  $store->add('added_id');
```

```
7     $store→remove('removed_id');
8
9     // Make assertions ...
10    $store→assertAdded('added_id');
11    $store→assertRemoved('removed_id');
12
13    $store→assertNotAdded('other_file_id');
14    $store→assertNotRemoved('other_file_id');
```

If a file is stored in the provider's [file storage](#) and added to a vector store in the same request, you may not know the file's provider ID. In this case, you can pass a closure to the `assertAdded` method to assert against the content of the added file:

```
1  use Laravel\Ai\Contracts\Files\StorableFile;
2  use Laravel\Ai\Files\Document;
3
4  $store→add(Document::fromString('Hello, World!', 'text/plain')→as('hello.txt')
5
6  $store→assertAdded(fn (StorableFile $file) => $file→name() === 'hello.txt')
7  $store→assertAdded(fn (StorableFile $file) => $file→content() === 'Hello, W
```

Events

The Laravel AI SDK dispatches a variety of [events](#), including:

`AddingFileToStore`

`AgentPrompted`

`AgentStreamed`

`AudioGenerated`

`CreatingStore`

`EmbeddingsGenerated`

`FileAddedToStore`

`FileDeleted`

`FileRemovedFromStore`

`FileStored`

`GeneratingAudio`

`GeneratingEmbeddings`

`GeneratingImage`

`GeneratingTranscription`

`ImageGenerated`

`InvokingTool`

`PromptingAgent`

`RemovingFileFromStore`

`Reranked`

`Reranking`

`StoreCreated`

`StoringFile`

`StreamingAgent`

`ToolInvoked`

`TranscriptionGenerated`

You can listen to any of these events to log or store AI SDK usage information.



Laravel is the most productive way to build, deploy, and monitor software.



© 2026 Laravel [Legal](#) [Status](#)

Products

Cloud

Forge

Nightwatch

Vapor

Nova

Packages

Cashier

Dusk

Horizon

Octane

Scout

Pennant

Pint

Sail

Sanctum

Socialite

Telescope

Pulse

Reverb

Echo

Resources

Documentation

Starter Kits

Release Notes

Blog

News

Community

Larabelles

Learn

Jobs

Caree

Partners

Curotec

Active Logic

Tighten

Kirschbaum

Vehikl

byte5

Redberry

Jump24

More Partners

Laravel